5.13 BROWSE DATA

Introduction

The StEPS Browse Data screen allows you to view one or two StEPS data sets in spreadsheet form, using the SAS data table view. This is a quick way of viewing some or all of the variables in a data set when you want to look up information quickly, but do not want to interactively edit the data. This screen also allows you to perform simple queries on the data set. The Browse Data screen is read only; any corrections to the data must be made in the interactive Review and Correction module.

Specifically, the StEPS Browse Data screen allows you to:

- ➤ View StEPS data in a spreadsheet layout.
- ➤ Compare two StEPS data sets side-by-side on one screen.
- ➤ Quickly search for data in a specified data set using scroll bars, the SAS Standard Where Clause, or by entering queries directly in the Browse screen.
- ➤ View summary statistics for the data set, including the total number of observations and the total number of variables in the data set.

Accessing the Screen(s)

There are two ways to access the Browse Data screen: via the Tools module or via the WhoamI screen:

1. To access the Browse Data screen via the Tools module:

   - **!** Click on the TOOLS button from the StEPS Main Menu.
   - **!** Click on the BROWSE DATA button from the Tools Menu to display a list of options.
   - **!** Select the “Browse One Data Set” option to display the screen shown in Figure 5.13.a. Select the “Browse Two Data Sets” option to invoke the screen shown in Figure 5.13.b.

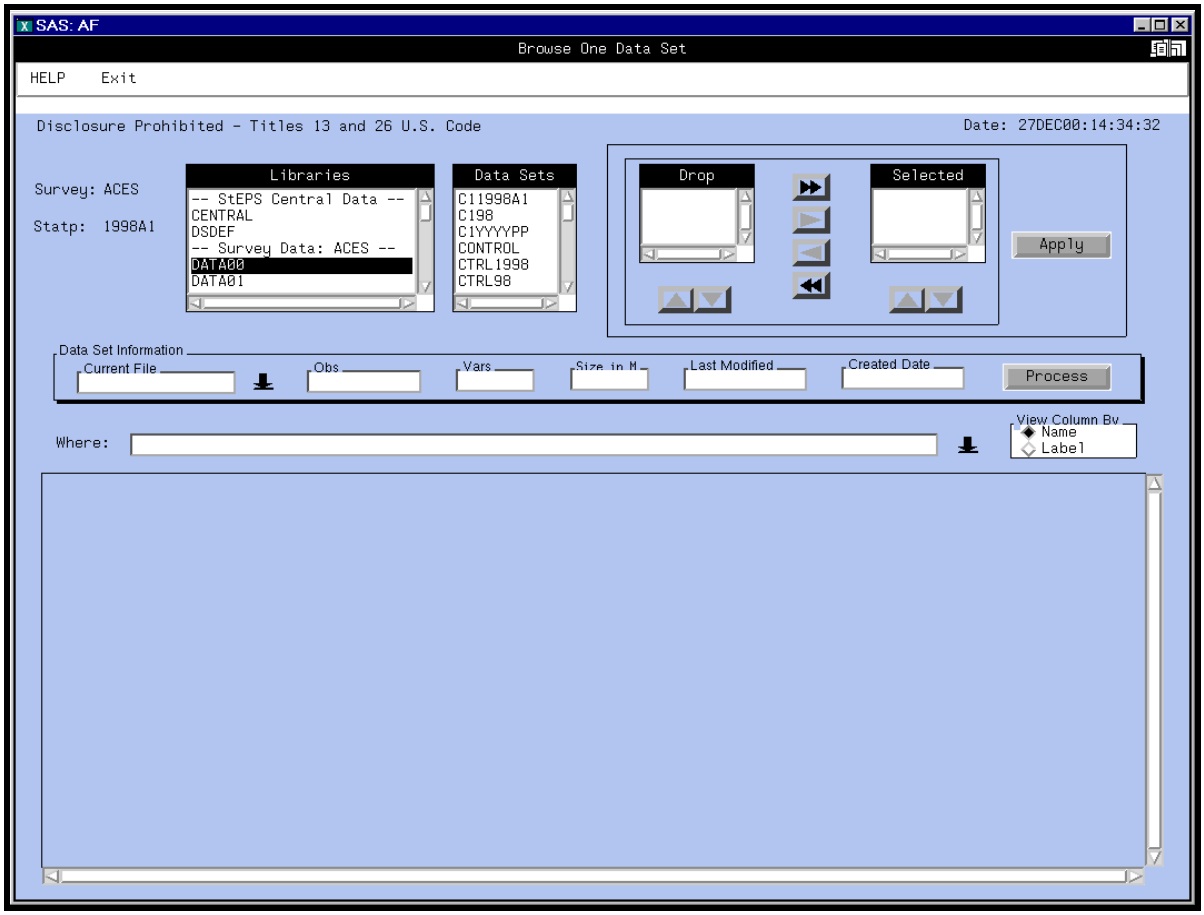


Figure 5.13.a Browse One Data Set screen

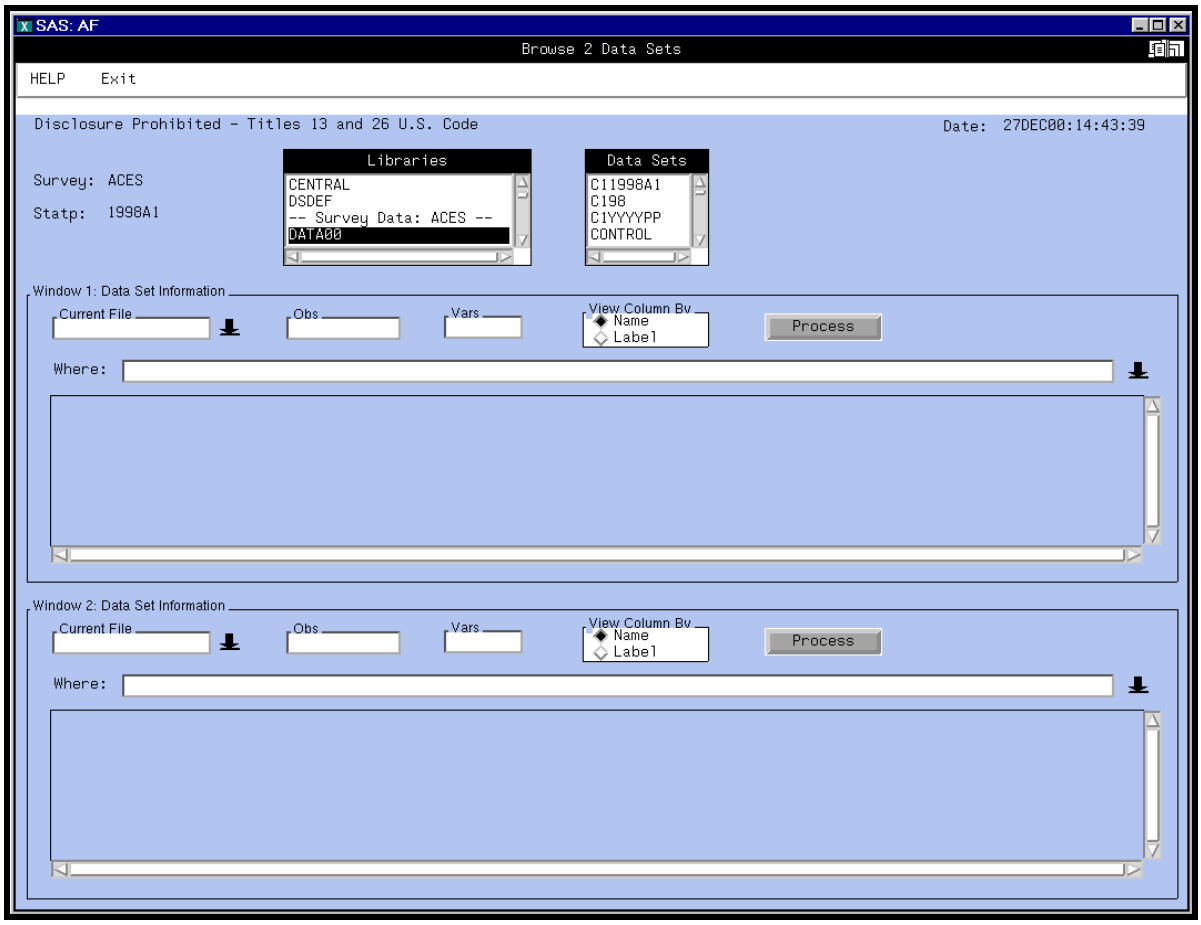


Figure 5.13.b Browse Two Data Sets screen

2. To access the Browse Data Screen via the “WhoamI” screen:

   ! Press the F7 function key (from anywhere within StEPS) to access the “WhoamI” screen (see Figure 5.13.d):

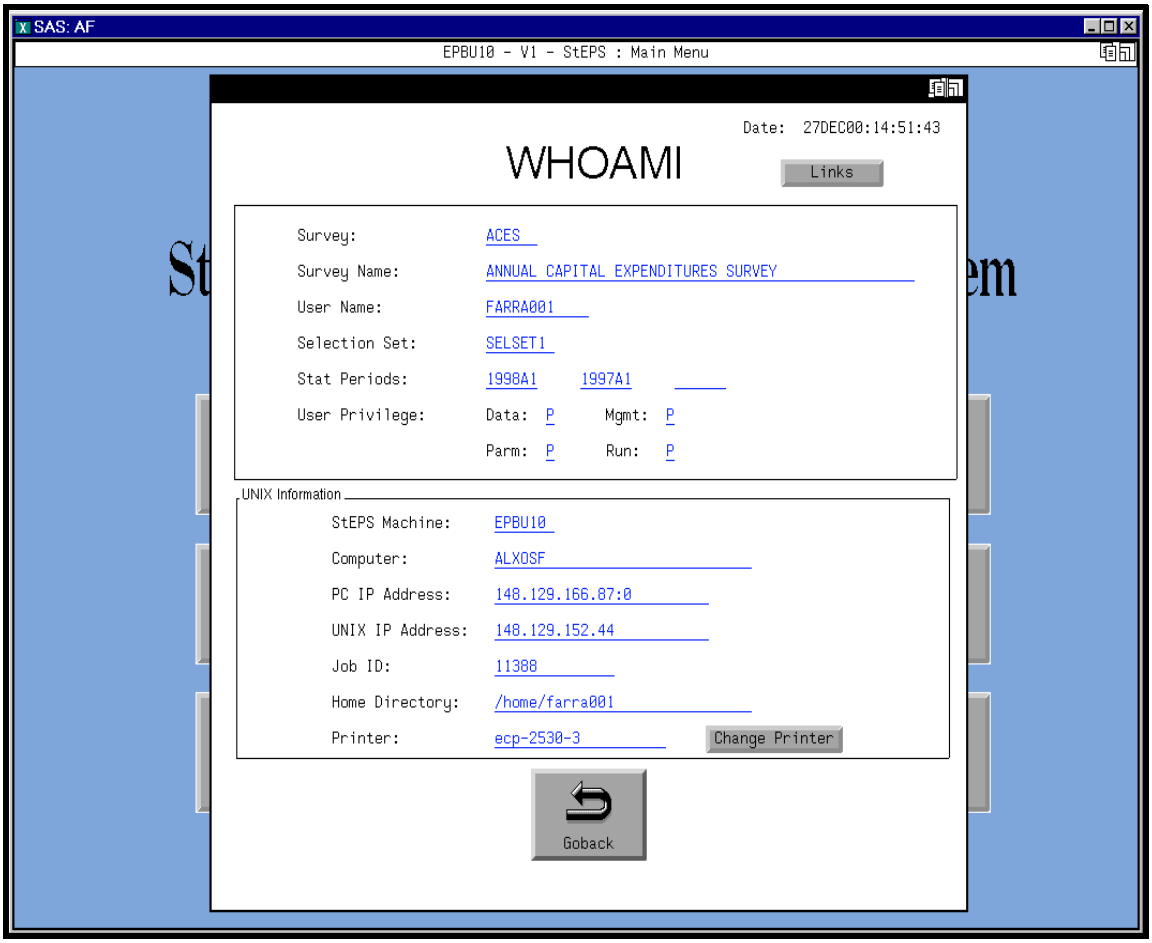


Figure 5.13.c "WhoamI" screen with links button, leading to Browse Data screens

! Click on the "Links" button in the upper right-hand corner of the "WhoamI" screen.

! Select either the "Browse One Data Set" or "Browse Two Data Sets" option from the pop-up menu (see Figure 5.13.d below) to display the desired screen illustrated in Figure 5.13.a or 5.13.b.

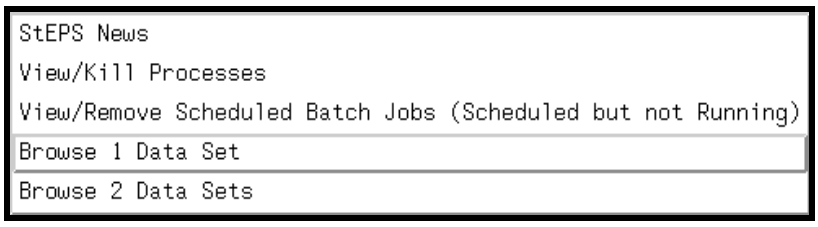


Figure 5.13.d Pop-up Picklist from "WhoamI" screen to access Browse Data Screen

Screen Features

Survey

- Survey identifier of survey currently being processed.
- Field not correctable.
- To change the survey, you must access the Survey Selection screen from USER SETUP (See Chapter 1.1).

Statp

- Statistical period of the survey being processed.
- Field not correctable.
- To change the stat period, you must access the Survey Selection screen from USER SETUP (See Chapter 1.1).

Libraries Window

- Displays the libraries from which you may select a data set to view.
- Select a library by clicking on the library name in the pick list.

Data Sets Window

- Displays the data sets contained in the library specified in the "Libraries Window".

Drop and Selected Windows

- The "Selected" window displays a list of variables to be included in the data set display. When a data set is initially specified, the default for the "Selected" window is ALL variables in the data set.
- To "remove" a variable from the "Selected" list, use the ◂◂ button.
- Move variables between windows using the • and ƒ buttons.
- Once all drop and select choices have been made, click on the "Apply" button to display the data set using only those variables specified.

Data Set Information

The following information is displayed in the "Data Set Information" section:

Current file	Name of file currently displayed on the screen (Note: Each file opened during the current session will be added to the picklist, accessible by clicking on the down arrow.)
Obs	Number of observations (records) in the data set.
Vars	Number of variables in the data set.
Size in M	Size, in megabytes, of data set currently displayed.
Last Modified	Date data set was last modified.
Created Date	Date data set was created.
Process	Click this button to access additional information about the data set, including:

1. Contents of the data set	Displays a SAS PROC CONTENTS of the data set
2. Search/Extract Data	Accesses a separate screen which allows you to subset the data set using a where clause, sort the output, save the output to a permanent data set, or save the data as an ID selection set. (See Chapter 5.10 for more information).

Where Clause Field

Subset the data set by entering a where clause directly in the where clause field or by invoking the StEPS Standard Where Clause. You can do the latter by 1) positioning the cursor in the data table and right clicking with the mouse or by 2) clicking on the "Process" button and choosing the "StEPS Standard Where Clause option".

See Section 5.13.3, below for detailed information on creating and entering where clauses.

"View Column By" Option

- You may alter the column labels that are displayed in the data table.
- The default setting is to display the variable "Name" (i.e. ID, PCFLG, etc.) as it exists in the data set.
- The alternative setting is "Label" which replaces the variable name with the descriptive label for that variable, as defined in the Item Data Dictionary.

5.13.1 BROWSING ONE DATA SET

To browse one data set:

1. Select a library from the "Libraries" box by clicking on the appropriate name of the library in the list. Available libraries include the following:

   - StEPS Central Data

     - CENTRAL — Data and parameter files that apply across StEPS; not specific to survey
     - DSDEF — Empty data set structures

   - Survey Data <Survey Name>

     - DATA00 — Current stat period data files
     - DATA01 — Prior stat period data files
     - DATA02 — Prior-prior stat period data files
     - SURVLIB — Files that define survey characteristics, such as forms, users, etc.
     - DATALIB — Survey data files that are not stat period specific
     - PARMLIB — Survey parameter files related to processes and modules
     - PARM00 — Parameter files specific to the current stat period
     - PARM01 — Parameter files specific to the prior stat period
     - PARM02 — Parameter files specific to the prior-prior stat period
     - ARCH00 — Archived data files specific to the current stat period
     - ARCH01 — Archived data files specific to the prior stat period
     - ARCH02 — Archived data files specific to the prior-prior stat period
     - STPSUSER — Data files related to the specific setup and defaults of the user
     - SELECT — User-created selection sets

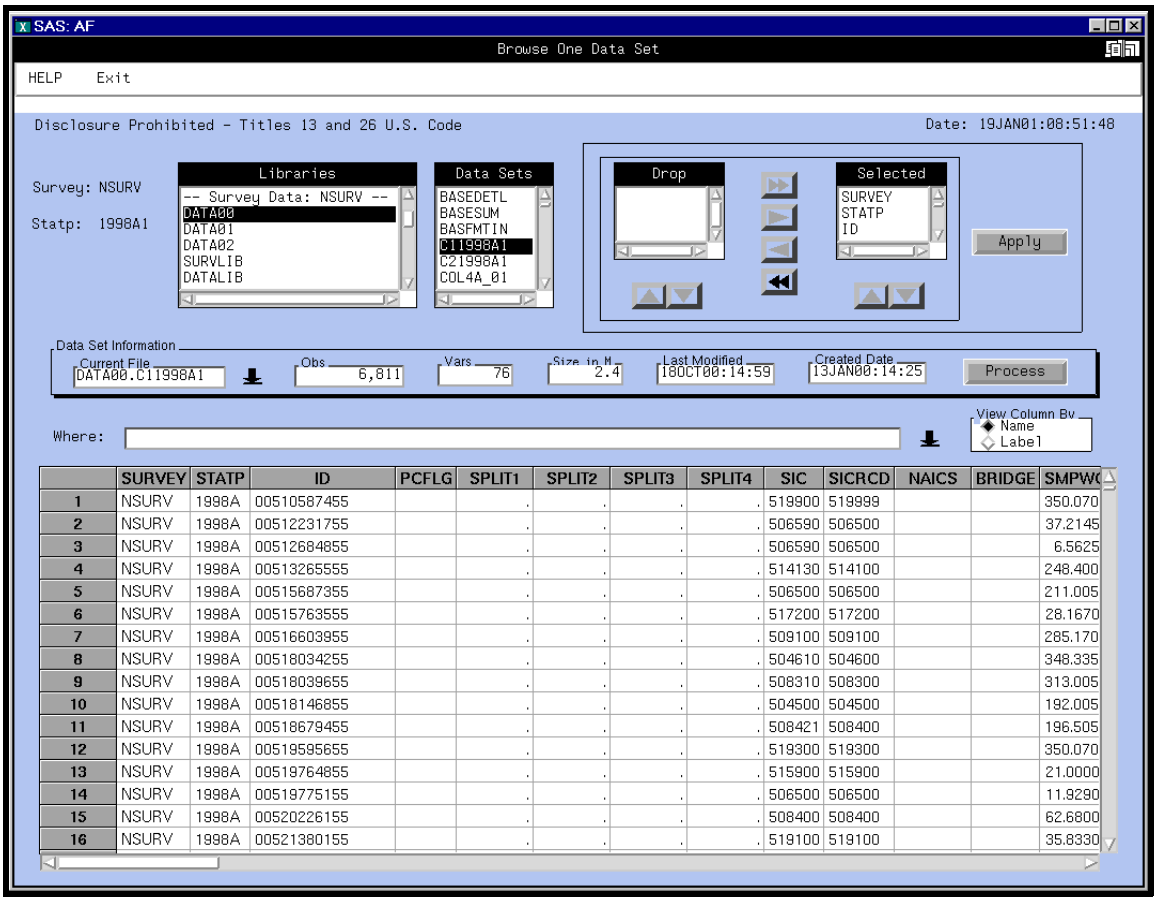
   - Temporary Data

     - USERTEMP — Temporary data files created by the user
     - WORK — Temporary data files created by StEPS

   - USER Data

     - USERLIB — Survey specific data, edit rejects, etc.

2. Select the data set you wish to view by clicking on the data set name within the “Data Sets” window. (The data sets available for display will depend upon the library you selected in the “Libraries” window.) The display should now look something like Figure 5.13.1.1:



	SURVEY	STATP	ID	PCFLG	SPLIT1	SPLIT2	SPLIT3	SPLIT4	SIC	SICRCD	NAICS	BRIDGE	SMPW
1	NSURV	1998A	00510587455		.	.	.	.	519900	519999			350.070
2	NSURV	1998A	00512231755		.	.	.	.	506590	506500			37.2145
3	NSURV	1998A	00512684855		.	.	.	.	506590	506500			6.5625
4	NSURV	1998A	00513265555		.	.	.	.	514130	514100			248.400
5	NSURV	1998A	00515687355		.	.	.	.	506500	506500			211.005
6	NSURV	1998A	00515763555		.	.	.	.	517200	517200			28.1670
7	NSURV	1998A	00516603955		.	.	.	.	509100	509100			285.170
8	NSURV	1998A	00518034255		.	.	.	.	504610	504600			348.335
9	NSURV	1998A	00518039655		.	.	.	.	508310	508300			313.005
10	NSURV	1998A	00518146855		.	.	.	.	504500	504500			192.005
11	NSURV	1998A	00518679455		.	.	.	.	508421	508400			196.505
12	NSURV	1998A	00519595655		.	.	.	.	519300	519300			350.070
13	NSURV	1998A	00519764855		.	.	.	.	515900	515900			21.0000
14	NSURV	1998A	00519775155		.	.	.	.	506500	506500			11.9290
15	NSURV	1998A	00520226155		.	.	.	.	508400	508400			62.6800
16	NSURV	1998A	00521380155		.	.	.	.	519100	519100			35.8330

Figure 5.13.1a Browse Data Screen with Data Set DATA00/C11998A1 Displayed

3. Note the summary statistics displayed for this data set: There are 6,811 observations (records/rows) and 76 variables (columns) in the file. In addition to this information, we can also see that the file is 2.4 megabytes in size and was created on 18 October 2000.

4. You may obtain summary information on one column of your data by performing a PROC MEANS from the data table. For example, you could use this feature to obtain summary statistics from the EDDATA column of your survey's item file. Furthermore, using the subsetting features described above, you can obtain summary statistics on just a portion of your items, such as the mean of all the eddata values for ID 11111111111 or all the rpdata values for item CSAL. To access this feature, you need to **left**-click your mouse within a cell for a numeric data column, such as rpdata or eddata from the item file (this feature will not work on character or date fields). Once you have left-clicked in the data field, a pop-up window will appear giving you several options. One of those options will be "Summarize the selected column". If you single-click on this option, StEPS will perform a PROC MEANS for you and the following output should appear in the SAS output window:

Summary of data set: DATA00.IT1998A1 12:49 Wednesday, August 15, 2001

The MEANS Procedure

Analysis Variable : EDDATA

N	N Miss	Minimum	Maximum	Mean	Sum	Std Error
506870	11744	-9.298877E12	1E13	384728369	1.9500727E14	56286078.68

Figure 5.10.5: Example of output from the PROC MEANS procedure.

- ! This output is the same as what you would receive if you ran a PROC MEANS within SAS (outside of StEPS). The information provided in this output is as follows:

    1. **N** = the number of observations in the data set on which the PROC MEANS was performed. In the example of figure 5.10.5, there are 506870 observations in the NSURV item file for 1998a1. If we had created a subset of the item file before performing the PROC MEANS, this value would have been smaller.

    2. **N Miss** = the number of missing values in the data column. In our example, there were 11744 missing values within eddata.

    3. **Minimum** = the smallest data value in the data column, in this case a huge negative number (-9.298877×10^{12})!

    4. **Maximum** = the largest data value in the data column.

    5. **Mean** = the mean value in the data column.

    6. **Sum** = the sum of all the values in the data column.

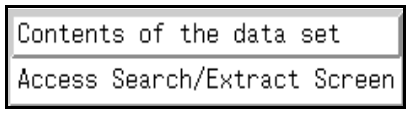
    7. **Std Error** = the standard error for the data column.

- ! Remember, the summary values you calculate will be for only one column in your data table. Thus, if you run a PROC MEANS on the eddata column, you will need to run a second PROC MEANS if you want to obtain summary values on the rpdata column.

5. Another nice feature is the ability to bookmark a particular row in the data table, making it easy to return to that row once you have scrolled on through the data table. To access this feature, left-click any data cell *within the row you are interested in.* A pop-up window will appear with the option "Set Bookmark: row #". For example, if you are interested in row 34 of the data table, you can left click any data cell within row 34, then click on the "Set Bookmark: Row 34" that appears on the pop-up window. Row 34 is now bookmarked and will remain so until you select another bookmark. If you scroll down to row 24592 and left click on any data cell, you will see two options in the pop-up window: One that says "Set New Bookmark: Row 24592" and another that says "Go To Bookmark: Row 34". If you select the "Go To" option, you will be returned to Row 34 of the data table. If you select the "Set New" option, Row 24592 will be the new bookmarked row. Row 34 will no longer be bookmarked. Thus, you may only set one bookmark at a time.

USING THE "PROCESS" FEATURE TO EXAMINE A DATASET:

The "Process" feature allows you to obtain additional information about the data set you are browsing. Clicking on the "Process" button pops-up the following menu of options:



You may access either of these options by clicking on it.

OPTION 1: CONTENTS OF THE DATASET

1. Select "Contents of the data set" to display the following:

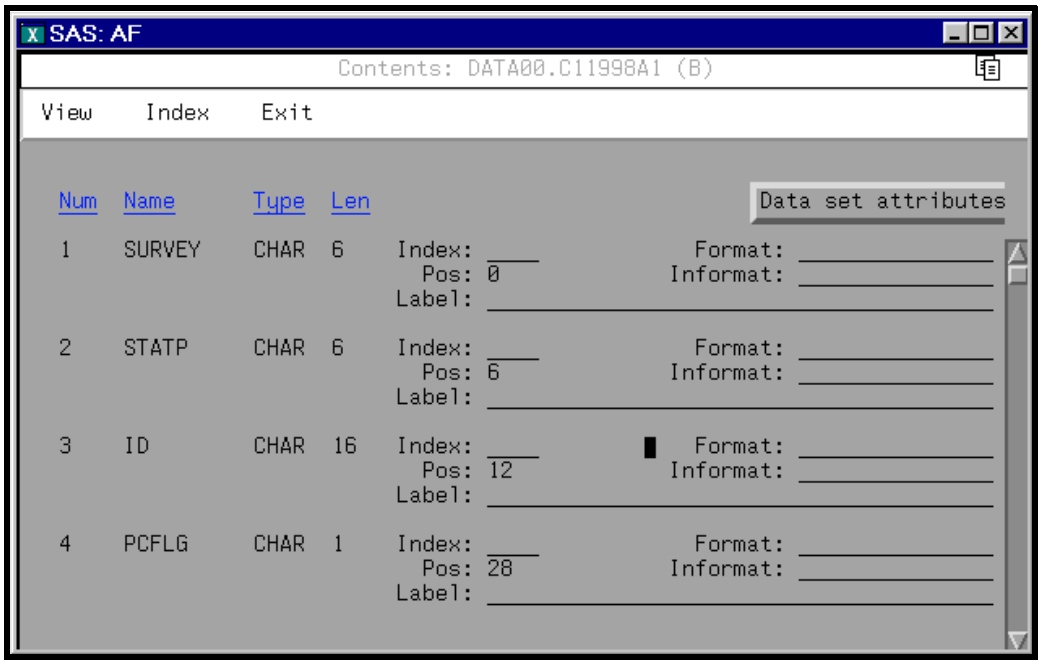


Figure 5.13.1b - Contents screen

2. Like a PROC CONTENTS in SAS, this screen provides structural information about the file itself. Specific information includes the following:

   ! NAME — Name of each variable in the data set. By using the "VIEW" p-menu, you can change the order in which the variable names are displayed: by name or by order:

1. **Sort by Name** displays the variable names in alphabetical order.
2. **Sort by Order** displays the variable names in the order in which they appear in the file.

- **!** TYPE — Variable type (i.e., character, numeric)
- **!** LENGTH — Length of the variable
- **!** INDEX — Indicates whether the variable is an index variable.
- **!** POSITION — Position of the variable within the file (the first variable in the file is at position 0).
- **!** Any formats or informats associated with the variable (e.g., date formats).

3. When you are finished reviewing the information in this screen, return to the Browse Data screen by using the EXIT p-menu option.

OPTION 2: SEARCH/ EXTRACT DATA USING WHERE CLAUSE SCREEN:

See Chapter 5.10 for detailed information on using the “Search/Extract” screen.

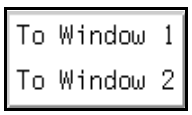
5.13.2 BROWSING TWO DATA SETS

To browse two data sets at the same time:

1. Select a library for the first data set from the “Libraries” box by clicking on the name of the appropriate library in the list. (See Figure 5.13.b)

2. Select the data set you wish to view by clicking on the data set name within the “Data Sets” window.

3. A pop-up window will display, prompting you to direct the selected data set to “window 1" or “window 2". Click on the window 1 option. The data set selected should now appear in window 1.



4. Select a library for the second data set from the “Libraries” box by clicking on the appropriate name of the library in the list.

5. Select the data set you wish to view by clicking on the data set name within the “Data Sets” window.

6. Select the “window 2" option from the pop-up window. You should now see something similar to Figure 5.13.2.1 below:

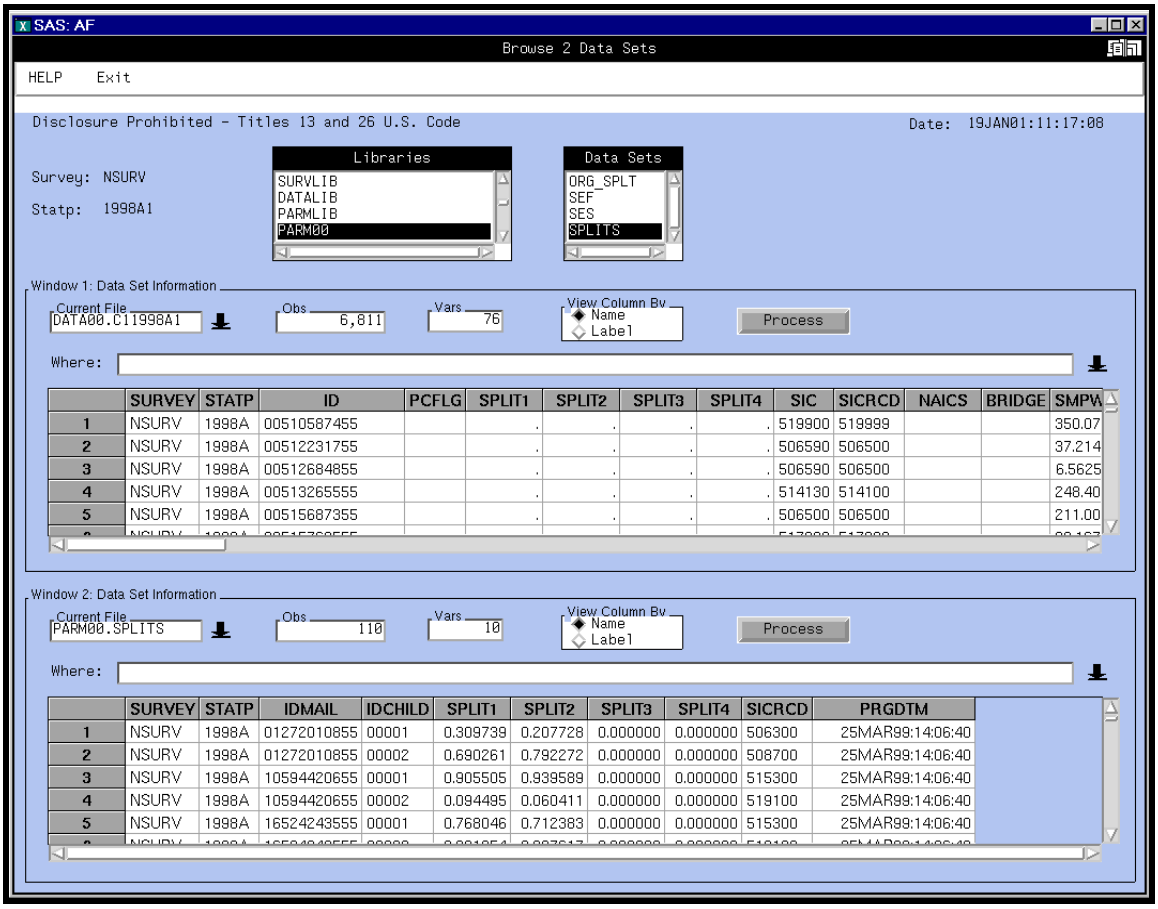


Figure 5.13.2 Browse Two Data Sets

NOTE: These two data sets are from two different libraries; you are not restricted to viewing two data sets from the same library. Additionally, each display window (1 & 2) provides summary statistics for the data set displayed. This includes the total number of observations and the total number of variables in each data set. You may also perform separate queries on the two data sets.

5.13.3 SUBSETTING DATA USING A WHERE CLAUSE:

You may reduce the number of records included in the display window using “where” clauses. You may define and run a where clause in one of two ways: 1) by invoking the SAS Standard Where Clause or 2) by typing the where clause directly in the “Where:” field.

15.13.3.1 USING THE SAS STANDARD WHERE CLAUSE

Invoke the SAS Standard Where Clause by 1) right-clicking within the data table, or by 2) selecting the down arrow next to the “Where:” field and then clicking on the Where Clause option from the pop-up menu. Choosing this option will access the following StEPS standard “where clause” screen:

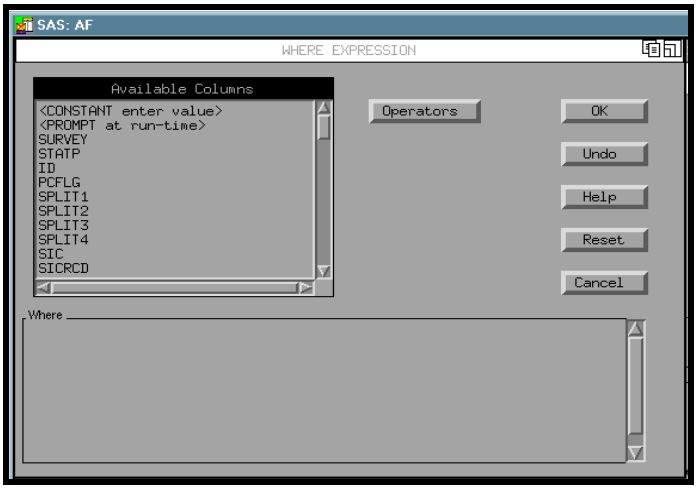


Figure 5.13.3a SAS Standard Where Clause Screen

1. To create a “where clause”, you must select a variable (from the “Available Columns” box), choose an operator, and select a constant. For example, if you wanted to limit the data to records where “NAICS = 503210" (‘NAICS’ is the variable, ‘EQ’ is the operator, and ‘503210' is the constant), you would do the following:

   - Select “NAICS” from the “Available Columns” box
   - Click on “Operators” to bring up a pick list of available operators (i.e., EQ, NE, GT, GE, LT, LE).
   - Click on “EQ” (equal to)
   - Click on “<CONSTANT enter value>” from the “Available Columns” box.
   - Another box will display. Enter the constant “503210" in the first box and press <ENTER> OR select “LOOKUP” to display a pick list of valid NAICS codes, from

which to choose.

NOTE: Click on the HELP button to access help information on using this screen.

2. You may choose a single variable or a combination of variables and operators to build your where clause. The variables listed in the "Available Columns" box are from the data set(s) displayed on your screen.

   NOTE: The dates listed in "LOOKUP" for control variables with date formats (i.e., check-in, extension) are displayed in a SAS format. The SAS format is for programming purposes and is not a readable format. To enter a date in the StEPS standard where clause screen for such variables, select the "CONSTANT" option and enter a date using the following format: 'DDmonthYY'd (where 'DD' = day and 'YY' = year) Example: January 13, 1999 = '13jan99'd)

3. As you build your "where clause", it will be displayed in the "Where" box at the bottom of the screen (as well as in the "Where:" field in the Browse Data screen.) Choose a button from the right side of the screen to do the following:

OK	Submit the where clause and adjust the data table display
UNDO	Remove the last variable, operator, or constant from the where clause
HELP	Access HELP information on using this screen
RESET	Clear your current where clause
CANCEL	Do not submit the where clause; return to the Browse Data Screen

4. Once the data table has been subsetted, you may revert to the display of all records in the entire data set by right-clicking in the data table and selecting the "Where Clear" option.

15.13.3.2 DEFINING A WHERE CLAUSE IN THE "WHERE:" FIELD

1. Type the variable(s), operator(s), and constant(s) into the window. There are no pick lists available from which to select. The where clause is not case sensitive, but you must observe SAS syntax. There is a "Syntax Check" feature available (see Figure 5.13.b)

2. Click on the down arrow to the right of the "Where:" field. This will pop-up the following menu:

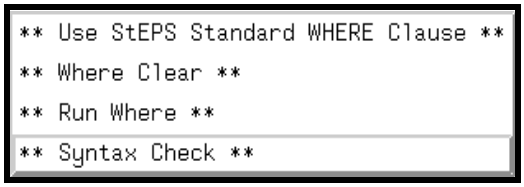


Figure 5.13.3b Where Clause Options

3. This menu allows you to do several things.

   - Invoke the SAS Standard Where Clause window described in Option 1 above. This is a good choice if you are unsure of the syntax for typing your where clause.
   - Clear the where clause you have just typed and start over.
   - Run the where clause to subset the data in the data display.
   - Check the syntax of the where clause you have entered before running it.

   Select the option of your choice by clicking on it.

4. Each where clause you create and run is saved in a temporary data file. These where clauses are then available to you as menu options (see Figure 5.13.3c below), accessible by clicking on the down arrow in the where clause field. These stored where clauses are lost once you exit the Browse Data screen.

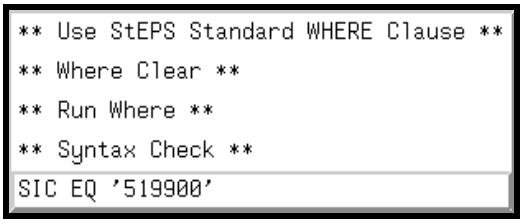


Figure 5.13.3c Stored Where clause

P-Menus

P-Menu	Options	Function
HELP	Browse Data Help (F1) WhoamI (F7)	Display HELP information on using the Browse Data screens. Display user default and systems information
EXIT	Exit (F3)	Exit to previous screen.